Local Convex Hull support and boundary estimation

Catherine Aaron, Olivier Bodart

Local Convex Hull support and boundary estimation

C. Aaron*, O. Bodart†

December 18, 2014

Abstract

In this paper we introduce a new estimator for the support of a multivariate density. It is defined as a union of convex hulls of observations contained in balls of fixed radius. We study the asymptotic behavior of this "local convex hull" for the estimation of the support and its boundary. When the support is smooth enough, the proposed estimator is proved to be, eventually almost surely, homeomorphic to the support. Numerical simulations on both simulated and real data illustrate the performance of our estimator.

Key Words: Convex-Hull, polyhedron, support estimation, topological data analysis, geometric inference.

AMS Classification : 62G05, 62G20, 62H99

1 Introduction

Let $\mathcal{X}_n = \{X_1, \ldots, X_n\}$ be a set of n independent and identically distributed (i.i.d) random variables with probability density f defined on $\mathbb{R}^d$. Let

$$S = \overline{\{x \in \mathbb{R}^d, f(x) > 0\}} \subset \mathbb{R}^d,$$

be the support of the probability density f, and

$$\partial S = S \setminus \mathring{S},$$

*Laboratoire de Mathématiques CNRS (UMR 6620) Université Blaise Pascal (Clermont-Ferrand 2) 63177 Aubière cedex, France (Catherine.Aaron@math.univ-bpclermont.fr)

†Laboratoire de Mathématiques CNRS (UMR 6620) Université Blaise Pascal (Clermont-Ferrand 2) 63177 Aubière cedex, France (Olivier.Bodart@math.univ-bpclermont.fr)

be its topological boundary (where $\overline{A}$ denotes the closure of the set A and $\mathring{A}$ its interior). We aim at building an estimator of S and ∂S based on the set of observations $\mathcal{X}_n$. Therefore, for a given estimator $\hat{S}_n$ of the support S, the following errors can be studied:

1. the measure of the symmetric difference:

$$|\hat{S}_n \Delta S| = |(\hat{S}_n \cap S^c) \cup (\hat{S}_n^c \cap S)|,$$

   (where A^c denotes the complementary of the set A and $|A|$ the Lebesgue measure of the set A);

2. the Hausdorff distance between $\hat{S}_n$ and S:

$$d_H(\hat{S}_n, S) = \max(\max_{a \in \hat{S}_n}(\min_{b \in S} \|a - b\|), \max_{b \in S}(\min_{a \in \hat{S}_n} \|a - b\|)),$$

   where $\|a - b\|$ denotes the Euclidean distance between a and b.

The first criterion is the most commonly used, but when it comes to evaluate the quality of the estimation of ∂S, only the Hausdorff distance $d_H(\partial \hat{S}_n, \partial S)$ between $\partial \hat{S}_n$ and ∂S is relevant.

The most intuitive and simple way to estimate S was introduced by Devroye and Wise (see [9] or [13]). They defined the following estimator:

$$\hat{S}_r = \bigcup_1^n \mathcal{B}(X_i, r), \tag{1}$$

where $\mathcal{B}(X, r)$ denotes the closed ball centered in x and of radius $r > 0$. The parameter r being set, building the estimator $\hat{S}_r$ is of low computational cost. Its properties have been widely studied for the estimation of S and ∂S. In [13], for a sequence of radiuses $r_n \to 0$ such that $nr_n^d \to \infty$, it has been proved to be universally consistent with respect to the symmetric difference. In [1], assuming standardness on S and f (see Definition 5 below), a rate of convergence of order $(\ln n/n)^{1/d}$ in terms of symmetric difference is established. With additional assumptions on f, central limit type theorems type are obtained in [3]. In [12], the authors study the estimation of the boundary of S in terms of the Haussdorff distance. Other estimators have also been proposed by various authors. In [18], the authors study the case when S is a subset of the unit square in $\mathbb{R}^2$. When f is an $\alpha-$decreasing function (see Definition 8 below), a piecewise polynomial estimator of ∂S is built and proved to have an optimal rate (which depends on α and the regularity of ∂S).

When S is a convex set, the convex hull of the observations is known to be a natural estimator of S (see e.g. [2, 24, 26, 14, 16]). This idea can be generalized to the non-convex case. In [25], the $r-$convex hull of the set of observations is studied as an estimator of S and ∂S. Assuming regularity on ∂S and that the probability density is bounded from below, it has the same convergence rate as the one obtain in [14] for a (similarly smooth) convex support, but with a weaker assumption on the support shape. Characterizing the $r-$convex hull via an erasing process, one can easily infer that the estimator can degenerate when S is a manifold of dimension $d' < d$. In [10], the authors build an analogous estimator with a different erasing process. This allows to weaken the assumptions on S. However the estimator presents the same degeneracy drawback.

In this work, we also propose a generalization of the convex hull, defined as follows:

$$\hat{H}_r = \bigcup_{X \in \mathcal{X}_n} \mathcal{H}(\mathcal{B}(X, r) \cap \mathcal{X}_n), \tag{2}$$

where $\mathcal{H}(A)$ denotes the convex hull of the set A.

Let us note that the first idea of using a local convex hull estimator has been introduced in [17], using nearest-neighbors instead of fixed radius. This estimator has good performances in different applications as home-range estimation, zoology and ecology (see e.g. [21], [19] or [22]). However, to our knowledge, the mathematical properties of this estimator have not been investigated so far.

In this paper, we study the estimator $\hat{H}_r$ defined by (2) as an estimator of the support S, and $\partial\hat{H}_r$ as an estimator of ∂S. Under general assumptions $\hat{H}_r$ (resp. $\partial\hat{H}_r$) will be proved to be consistent in terms of the symmetric difference (resp. the Hausforff distance). Under assumptions on the probability density f and geometrical assumptions on S, convergence rates will be exhibited. Moreover, topological properties of $\hat{H}_r$ will also be given. More precisely, it will be proved to be eventually almost surely homeomorphic to S. The practical interest of such a property is illustrated in [7, 5]. Such a property is more frequently studied in the field of computational or discrete geometry than in statistics (see e.g. [15, 4, 6, 28, 8]).

The paper is organized as follows. Section 2 is devoted the presentation of the probabilistic and geometrical framework of the article, and of our main results. Theorem 1 deals with the consistency of $\hat{H}_r$. Theorem 2 and 3 give convergence rates. Finally, Theorem 4 studies the topological properties of our estimator. In Section 3, Theorems 1 and 2 are proved. Sections 4 and 5 are devoted to the proof of Theorems 3 and 4 respectively. The last section present numerical simulations.

2 General framework and main results

We will start this section with the setting of the geometric and probabilistic framework of this work. We will then state our results.

2.1 Notation and geometrical concepts

Throughout the paper, ω_d will denote the Lebesgue measure of the unit ball in $\mathbb{R}^d$, $d \geq 1$. The measure of a set $A \in \mathbb{R}^d$ will be denoted by $|A|$, and $\|x\|$ will denote the euclidean norm of $x \in \mathbb{R}^d$. Moreover, for a set $A \in \mathbb{R}^d$ and a positive real number ε, $A \oplus \varepsilon\mathcal{B}$ (reps. $A \ominus \varepsilon\mathcal{B}$) will denote the Minkowski sum (resp. difference) of A and balls of radius ε, that is

$$A \oplus \varepsilon\mathcal{B} = \bigcup_{a \in A} \mathcal{B}(a, \varepsilon); \qquad A \ominus \varepsilon\mathcal{B} = \{a \in A, \mathcal{B}(a, \varepsilon) \subset A\}.$$

Finally, the Hausdorff distance between two sets A and B in $\mathbb{R}^d$ is defined as follows:

$$d_H(A, B) = \inf\{\varepsilon, A \subset B \oplus \varepsilon\mathcal{B}, B \subset A \oplus \varepsilon\mathcal{B}\}.$$

The main geometrical objects under consideration in this paper are manifolds in $\mathbb{R}^d$.

Definition 1. *Two topological spaces A and B are homeomorphic (denoted by $A \approx B$) if there exists a bijective map $\varphi : A \to B$ such that φ and φ^{-1} are continuous.*

Definition 2. *A bounded set $A \subset \mathbb{R}^d$ is a $k-$dimensional manifold (or more shortly a $k-$manifold) if, for each point $x \in A$, there exists a neighborhood U_x of this point such that $U_x \approx \mathbb{R}^k$ or $U_x \approx \{(x_1, \ldots, x_k) \in \mathbb{R}^k, x_1 \geq 0\}$. Its boundary ∂A is the union of the points x which don't admit any neighborhood $U_x \approx \mathbb{R}^k$.*

A $k-$manifold is thus locally homeomorphic to a plane of dimension k, and its boundary (if it exists) is homeomorphic to a half plane. It is worth noticing that the definition of the boundary of a manifold coincides with the classical one (i.e. $\partial A = \overline{A} \setminus \mathring{A}$) when the natural topology induced by A is chosen. Moreover we have the classical results:

Proposition 1. *If $A \in \mathbb{R}^d$ is a $k-$manifold with non-empty boundary ∂A then ∂A is a $(k-1)-$manifold.*

Proposition 2. *If A is a compact $d-$manifold in $\mathbb{R}^d$, then it has a non empty boundary and for all $x \in A$, $y \in A^c$ the line segment $[x, y]$ intersects the boundary: $[x, y] \cap \partial A \neq \emptyset$.*

Let us notice that Proposition 2 does not hold for a $k-$manifold in $\mathbb{R}^d$ with $k < d$. Definition 2 is the topological definition of a manifold. The notion of differentiable manifold (or more generally of a manifold of class $\mathcal{C}^p$) is also relevant. It is essentially related to the regularity of the local homeomorphisms featured in the topological definition. For more details about this question, and about manifolds and differential geometry we refer to [20]. The concept of homeomorphic sets is also relevant to another aspect of this work: if the estimator of the support S is homeomorphic to S, then its topological properties are preserved by the estimation process.

A $k-$manifold is of topological dimension k; in this work, another dimension parameter is also used.

Definition 3. *Let $A \subset \mathbb{R}^d$ be a compact set, and let $N(A, \varepsilon)$ denote the minimal number of balls of radius ε that are needed to cover A. The Minkowski (or box-counting) dimension of A is defined, if it exists, by:*

$$\mathrm{Dim}_{Mink}(A) = \lim_{\varepsilon \to 0} -\frac{\ln N(A, \varepsilon)}{\ln \varepsilon}.$$

There exist sets which have no Minkowski dimension, which will not be considered in this paper. We will consider sets which have a Minkowski dimension that can be different from their topological dimension. Such sets can be rather pathological (consider for example a Koch snowflake in $\mathbb{R}^2$ which has Minkowski dimension 4/3), but in the case of smooth enough manifolds, the situation is rather simple.

Proposition 3. *If $A \subset \mathbb{R}^d$ is a $k-$manifold of class $\mathcal{C}^1$, with $k \leq d$ then $\mathrm{Dim}_{Mink}(A) = k$. Moreover there exists λ_A such that $N(A, \varepsilon) \leq \lambda_A \varepsilon^{-k}$.*

Notice that a compact $d-$manifold in $\mathbb{R}^d$ is necessarily of class $\mathcal{C}^\infty$ (see [20]).

We now introduce definitions in concepts in a more direct relationship with our concerns.

Definition 4. *A sequence (A_n) of events in the same probability space is told to happen eventually almost surely, denoted by e.a.s. if the sequence of random variables 1_{A_n} satisfies : $1_{A_n} \xrightarrow{a.s.} 1$.*

In the sequel we will make use of the following classical result.

Proposition 4. *If (A_n) and (B_n) are two sequences of events such that (A_n) happens e.a.s and, for n large enough, $A_n \Rightarrow B_n$ then (B_n) happens e.a.s.*

One of our main results (Theorem 2) deals with density supports that have a regularity property called standardness and partial expandability which are defined as follows.

Definition 5. *A measure is said to be standard with respect to the Lebesgue measure if there exists $\lambda > 0$ and $\delta > 0$ such that: $\mathbb{P}_X(\mathcal{B}(x,\varepsilon)) \geq \delta|\mathcal{B}(x,\varepsilon)|$ for all $x \in S$ and $\varepsilon \in]0,\lambda]$.*

Definition 6. *A bounded Borel set $S \subset \mathbb{R}^d$ is said to be partly expandable if there exists constants $C_S \geq 1$ (called expandability constant) and $r_S > 0$ such that $d_H(\partial S, \partial S \oplus \varepsilon\mathcal{B}) \leq \varepsilon\, C_S$ for all $\varepsilon \in]0, r_S]$.*

The concept of standardness was first introduced in [11] . We refer to [12] for the study of partial expandability. Standardness is a geometrical and probabilistic property. However, partial expandability is a stronger geometrical assumption than standardness. Therefore, if both are assumed, only the probabilistic consequences of standardness remain significant. Mainly, making these two assumptions ensures that the probability density decays fast enough to 0, and prevents the existence of cusps in the boundary of the support. However, they are satisfied by a large class of nonsmooth supports. For example, a uniform probability measure defined inside a Koch snowflake in $\mathbb{R}^2$ is standard and its support is partially expandable. It also has to be remarked that, when the underlying probability density f is smooth enough, a direct consequence of standardness and partial expandability is that f possesses a positive uniform lower bound on its support. We aim at weakening such an assumption. This will lead to stronger geometrical assumptions on the shape of the support.

Definition 7. *Let S be a $d-$manifold in $\mathbb{R}^d$ with non empty boundary. We say that balls of radius $R_{out} > 0$ (resp. $R_{in} > 0$) roll freely outside (resp. inside) S if, for all $x \in \partial S$ there exists $O_x^- \in \mathbb{R}^d$ (resp. $O_x^+ \in \mathbb{R}^d$) such that: $x \in \mathcal{B}(O_x^-, R_{out}) \subset \overline{S^c}$ (resp. $x \in \mathcal{B}(O_x^+, R_{in}) \subset \overline{S}$).*

This property has regularity consequences which will be used in the sequel.

Proposition 5. *Let S be a compact $d-$manifold in $\mathbb{R}^d$ such that balls of radius R_S roll freely inside and outside S. Then, for all $x \in \partial S$, there exists a unique inward pointing unit normal vector u_x and*

$$\forall (x,y) \in \partial S^2,\ ||u_y - u_x|| \leq \frac{1}{R_S}\,||x-y||.$$

Moreover ∂S is a $(d-1)-$dimensional manifold of class $\mathcal{C}^1$ and there exists a positive constant $\lambda_{\partial S}$ such that $N(\partial S, \varepsilon) \leq \lambda_{\partial S}\, \varepsilon^{k-1}$ (hence $\mathrm{Dim}_{Mink}(\partial S) = k-1$).

Proposition 6. *Let S be a compact $d-$manifold in $\mathbb{R}^d$ such that balls of radius R_S roll freely inside and outside S. Then S is partly expandable with expandability constant $C_S = 1$.*

The proof of Proposition 5 can be found in [27]. For Proposition 6, we refer to [12]. We will also use the following result.

Proposition 7. *Let S be a compact $d-$manifold in $\mathbb{R}^d$ such that balls of radius R_S roll freely inside and outside S. Then, for all x such that $d(x, \partial S) < R_S$, there exists a unique $x^* \in \partial S$ such that $d(x, x^*) = \min_{z \in \partial S} ||x - z||$. Moreover $x - x^*$ is collinear to u_{x^*}.*

Proof. Let x be a point such that $x \in S$ and $d(x, \partial S) = d < R_S$. If y_1 and y_2 are two points in ∂S such that $||x - y_1|| = ||x - y_2|| = d$ then $\mathcal{B}(x, d) \subset S$. The (inside and outside) rolling ball property implies that $O^+_{y_1}$, x, y_1 and $O^-_{y_1}$ are on the same line directed by u_{y_1}. Hence, the condition $d < R_S$ implies that $(\mathring{\mathcal{B}}(O^+_{y_1}, R_S))^c \cap \mathcal{B}(x, d) = \{y_1\}$. But, $y_2 \in \mathcal{B}(x, d)$ and $y_2 \notin \mathring{\mathcal{B}}(O^+_{y_1}, R_S))$ (because $y_2 \notin \mathring{S}$) so $y_2 = y_1 = x^*$ and x, $O^+_{x^*}$, $O^-_{x^*}$, and x^* are on the same line directed by u_{x^*}.

A symmetrical reasoning can be done when x belong to S^c and $d(x, \partial S) = d < R_S$. □

Remark: In the sequel, we will use the notations O^+_x, O^-_x and u_x that are introduced in Definition 7 and Proposition 5.

The rolling balls property being stronger than partial expandability and standardness, this will allow us to weaken the assumption on the probability distribution of the sample $\mathcal{X}_n$.

Definition 8. *A probability density f supported in $S \subset \mathbb{R}^d$ is said to be $\alpha-$quickly decreasing if there exists $\alpha \geq 0$ and $C_f > 0$ such that*

$$\forall x \in S,\ f(x) \geq C_f\, d(x, \partial S)^\alpha.$$

This assumption is indeed weaker than standardness. It is similar to the one introduced in [18].

Finally, when dealing with manifolds satisfying the rolling ball property, we will make use of the notion of tangent cylinder.

Definition 9. *Let u be a unit vector in $\mathbb{R}^d$, $x \in \mathbb{R}^d$, $r > 0$, and $h > 0$. The cylinder at point x in the direction u of radius r and height h is defined as follows:*

$$\mathcal{C}^u(x, r, h) = \{y, |\langle y - x, u\rangle| \leq h, ||y - x - \langle y - x, u\rangle u|| \leq r\}$$

The following result then holds.

Proposition 8. *Let x and x' be two points and u and u' be two unit vectors. Let us denote $||x - x'|| = \varepsilon_x$ and $||u - u'|| = \varepsilon_u$. We define:*

i) $e_1 = \varepsilon_x + 2(\sqrt{r^2 + h^2} + \varepsilon_x)\varepsilon_u$,

ii) $e_2 = \varepsilon_x + \sqrt{r^2 + h^2}\varepsilon_u$.

When ε_x and ε_u are small enough so that $e_1 \leq r$ and $e_2 \leq h$, we have:

$$\mathcal{C}^{u'}(x', r - e_1, h - e_2) \subset \mathcal{C}^u(x, r, h).$$

The proof is easy and left to the reader.

Definition 10. *Let S be a compact $d-$manifold in $\mathbb{R}^d$ such that balls of radius R_S roll freely inside and outside S. Let $x \in \partial S$, and u_x as in Proposition 5. For $r > 0$ and $h > 0$, the cylinder $\mathcal{C}(x, r, h) = \mathcal{C}^{u_x}(x, r, h)$ is called a tangent cylinder to ∂S.*

2.2 Main results

The first result we prove in this article is a universal consistency theorem.

Theorem 1. *Let $\mathcal{X}_n = \{X_1, \ldots, X_n\}$ be a set of i.i.d random observations in $\mathbb{R}^d$, $d \geq 1$, which distribution $\mathbb{P}_X$ is absolutely continuous with respect to the Lebesgue measure and supported in a compact $d-$manifold $S \in \mathbb{R}^d$ which boundary is such that $\mathrm{Dim}_{Mink}(\partial S) = d' < d$. Assume that there exists a sequence of radiuses (r_n) such that*

$$r_n \to 0 \text{ a.s. and } S \subset \hat{S}_{r_n/4} \text{ e.a.s.} \tag{3}$$

where $\hat{S}_{r_n/4}$ is the Devroye-Wise estimator defined by (1)*. The estimator $\hat{H}_{r_n}$ defined by* (2) *then satisfies:*

$$\forall u > 0, |\hat{H}_{r_n} \Delta S| / r_n^{d-d'-u} \to 0 \text{ a.s.}, \tag{4}$$

$$d_H(\partial \hat{H}_{r_n}, \partial S) \to 0 \text{ a.s.} \tag{5}$$

In [12], it is proved that the sequence $r_n = 4\, d_H(\mathcal{X}_n, S)$ fulfills assumption (3). However this sequence of radiuses is abstract and its convergence rate to 0 cannot be calculated. Therefore Theorem 1 provides us with no convergence rate of the estimators of S and ∂S. This is the aim of our two next results: thanks to additional assumptions on S and $\mathbb{P}_X$, we will exhibit explicit sequences (r_n) allowing to estimate the convergence rate in (4) and (5).

Theorem 2. *Let $\mathcal{X}_n = \{X_1, \dots, X_n\}$ be a set of i.i.d random observations in $\mathbb{R}^d$, $d \geq 1$, which distribution $\mathbb{P}_X$ is absolutely continuous with respect to the Lebesgue measure and supported in a compact $d-$manifold $S \in \mathbb{R}^d$ which boundary is such that $\mathrm{Dim}_{Mink}(\partial S) = d' < d$. Assume that S is partly expandable (with a expandability constant C_S) and that $\mathbb{P}_X$ is standard with respect to the Lebesgue measure. Consider the radius sequence:*

$$r_n = c \cdot \left(\frac{2\ln(n)}{\delta \omega_d n} \right)^{1/d} \quad \text{for some } c > 4, \tag{6}$$

where $\delta > 0$ is the constant appearing in Definition 5. Then the estimator $\hat{H}_{r_n}$ defined by (2) satisfies (4). Moreover we have

$$d_H(\partial S, \partial \hat{H}_{r_n}) \leq C_S \, r_n \ \text{e.a.s.} \tag{7}$$

Notice that (7) provides the classical rate of order $(\ln n/n)^{1/d}$ for the estimator of the boundary. The convergence rate given by (4) is weaker and depends on the Minkowski dimension of ∂S. However, assuming that ∂S is of class $\mathcal{C}^1$, one can easily obtain the same classical convergence rate for the symmetric difference $\hat{H}_{r_n} \Delta S$.

With a smoothness assumption on the boundary of S, and an adequate choice of the sequence of radiuses, we obtain a better convergence rate. More precisely we have the following Theorem.

Theorem 3. *Let $\mathcal{X}_n = \{X_1, \dots, X_n\}$ be a set of i.i.d random observations in $\mathbb{R}^d$, $d \geq 1$, which distribution $\mathbb{P}_X$ is absolutely continuous with respect to the Lebesgue measure and supported in a compact $d-$manifold $S \in \mathbb{R}^d$. Assume that the probability f density associated to $\mathbb{P}_X$ is $\alpha-$quickly decreasing and that there exists $R_S > 0$ such that balls of radius R_S roll freely inside and outside S. Let*

$$r_n = \lambda \left(\ln n/n\right)^{1/(d+1+2\alpha)}, \quad \lambda > 0. \tag{8}$$

We then have :

$$|\hat{H}_{r_n} \Delta S| \left(n/\ln n\right)^{2/(d+1+2\alpha)} \ \text{e.a.s. bounded}, \tag{9}$$

$$d_H(\partial \hat{H}_{r_n}, \partial S) \left(n/\ln n\right)^{2/(d+1+2\alpha)} \ \text{e.a.s. bounded}. \tag{10}$$

When $d = 2$, the convergence rate given here is the same as in [18] (our assumptions on the boundary smoothness being similar). In higher dimensions, when $\alpha = 0$, it is the same as the one obtain in [14] using the convex hull to estimate a (similarly smooth) convex support under the same assumption on the density. This suggests that our result might be optimal, although it is not proved here.

The same assumptions also yield the following result on the preservation of the topology of the support and its boundary.

Theorem 4. *Under the assumptions of Theorem 3 on S, $\mathbb{P}_X$ and the sequence (r_n), we have:*

$$\partial \hat{H}_{r_n} \text{ is homeomorphic to } \partial S \text{ e.a.s.}, \tag{11}$$

$$\hat{H}_{r_n} \text{ is homeomorphic to } S \text{ e.a.s.} \tag{12}$$

3 Proof of Theorems 1 and 2

3.1 Proof of Theorem 1

In a first step we are going to prove that

$$(\hat{H}_{r_n} \cap S^c) \cup (\hat{H}^c_{r_n} \cap S) \subset \partial S \oplus r_n \mathcal{B} \text{ e.a.s.} \tag{13}$$

Let us set $r = r_n$ fixed temporarily for the sake of clarity. From (1) and (2), we clearly have $\hat{H}_r \subset \hat{S}_r \subset S \oplus r\mathcal{B}$, hence $\hat{H}_r \cap S^c \subset (S \oplus r\mathcal{B}) \cap S^c$. Let $x \in (S \oplus r\mathcal{B}) \cap S^c$. In view of Proposition 2 there exists $y \in S$ such that $||x - y|| \leq r$ and the line segment $[x, y]$ intersects ∂S at some point z which is obviously such that $||x - z|| \leq r$. We thus have $(S \oplus r\mathcal{B}) \cap S^c \subset \partial S \oplus r\mathcal{B}$ that is

$$\hat{H}_r \cap S^c \subset \partial S \oplus r\mathcal{B}. \tag{14}$$

Next, we write

$$\hat{H}^c_r \cap S = (\hat{H}^c_r \cap (S \ominus \frac{r}{2}\mathcal{B})) \cup (\hat{H}^c_r \cap (S \setminus (S \ominus \frac{r}{2}\mathcal{B}))).$$

Let us first prove that $\hat{H}^c_r \cap (S \ominus \frac{r}{2}\mathcal{B}) = \emptyset$. Assume it is not the case, and let $x \in \hat{H}^c_r \cap (S \ominus \frac{r}{2}\mathcal{B})$. Then, necessarily, $x \notin \mathcal{H}(\mathcal{B}(x, r/2) \cap \mathcal{X}_n)$. Indeed, if $x \in \mathcal{H}(\mathcal{B}(x, r/2) \cap \mathcal{X}_n)$, then there exists at least one observation $X_i \in \mathcal{B}(x, r/2)$ and, as $\mathcal{B}(x, r/2) \subset \mathcal{B}(X_i, r)$, we have $x \in \mathcal{H}(\mathcal{B}(X_i, r) \cap \mathcal{X}_n)$ so that $x \in \hat{H}_r$. Now, since $x \notin \mathcal{H}(\mathcal{B}(x, r/2) \cap \mathcal{X}_n)$ there exists a unit vector u such that any observation $X_i \in \mathcal{X}_n \cap \mathcal{B}(x, r/2)$ satisfies $\langle X_i - x, u\rangle \leq 0$. Let $y = x + (r/4)u$ as depicted in Figure 1. We have $||x - y|| = r/4$, and, since $x \in (S \ominus \frac{r}{2}\mathcal{B})$, the inclusions $\mathcal{B}(y, r/4) \subset \mathcal{B}(x, r/2) \subset S$ hold. Hence $y \in S$. But we also have $\mathcal{B}(y, r/4) \cap \mathcal{X}_n = \emptyset$, that is $y \in \hat{S}^c_{r_n/4}$ which is impossible since, due to assumption (3), $S \cap \hat{S}^c_{r/4} = \emptyset$. Hence we have

$$\hat{H}^c_r \cap S = \hat{H}^c_r \cap S \setminus (S \ominus \frac{r}{2}\mathcal{B})) \subset S \setminus (S \ominus \frac{r}{2}\mathcal{B}).$$

Let $x \in S \setminus (S \ominus \frac{r}{2}\mathcal{B})$. Proposition 2 implies the existence of $y \in \mathcal{B}(x, r/2) \cap S^c \neq \emptyset$ such that the line segment $[x, y]$ intersects ∂S at some point z such that $||x - z|| \leq r/2$. Hence, we have

$$\hat{H}_r^c \cap (S \setminus (S \ominus \frac{r}{2}\mathcal{B})) \subset S \setminus (S \ominus \frac{r}{2}\mathcal{B}) \subset \partial S \oplus (r/2)\mathcal{B}. \tag{15}$$

Formula (13) is a direct consequence of (14), (15) and Proposition 4.

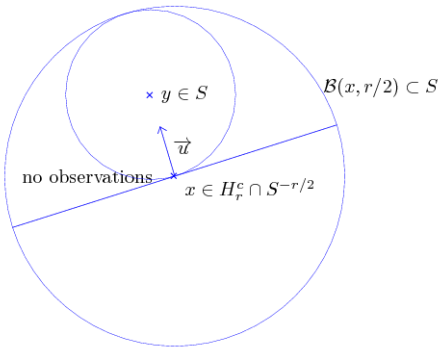


Figure 1: $x \in H_r^c \cap (S \ominus \frac{r}{2}\mathcal{B}) \Rightarrow \exists y \in S\,;\, \mathcal{B}(y, r/4) \cap \mathcal{X}_n = \emptyset$

Now, for fixed n, the parameter $N(\partial S, r_n)$ being as given in Definition 3, we have

$$\partial S \subset \bigcup_{i=1}^{N(\partial S, r_n)} \mathcal{B}(x_i, r_n),$$

that is, due to the triangle inequality,

$$\partial S \oplus r_n\mathcal{B} \subset \bigcup_{i=1}^{N(\partial S, r_n)} \mathcal{B}(x_i, 2r_n),$$

hence $|\partial S \oplus r_n\mathcal{B}| \leq N(\partial S, r_n)\omega_d 2^d r_n^d$.

Since $\mathrm{Dim}_{Mink}(\partial S) = d' < d$, in view of Definition 3 and assumption (3), we have $\ln N(\partial S, r_n)/\ln r_n \to -d'$ as $n \to \infty$, that is

$$\forall u > 0,\ \exists n_u;\ n \geq n_u \longrightarrow \ln N(\partial S, r_n)/\ln(r_n) \geq -d' - u/2.$$

Therefore, for any $u > 0$, we have $|\partial S \oplus r_n\mathcal{B}|/r_n^{d-d'-u} \leq \omega_d 2^d r_n^{u/2} \to 0$, which implies, in view of Proposition 4, that

$$\forall u > 0, \quad |\partial S \oplus r_n\mathcal{B}|/r_n^{d-d'-u} \to 0 \ \ e.a.s. \tag{16}$$

Combining (13) and (16), we obtain (4).

Let us now prove (5). First, we have

$$\max_{x\in\partial S} d(x,\partial\hat{H}_{r_n}) \to 0 \ a.s. \tag{17}$$

We will only sketch the proof of this convergence, since it is an adaptation to our case of the proof of Theorem 1 in [12]. Suppose (17) does not hold. Then, there exists $\varepsilon > 0$ and a subsequence of indexes (n_k) and a sequence (y_{n_k}) of points in ∂S such that $\mathbb{P}(d(y_{n_k}, \partial\hat{H}_{r_{n_k}}) > 2\varepsilon) \neq 0$. The boundary ∂S being compact, we can extract a subsequence, still denoted (y_{n_k}) for convenience, converging to $y \in \partial S$. The triangle inequality implies that $d(y, \partial\hat{H}_{r_{n_k}}) > \varepsilon$ for k large enough, that is $\mathcal{B}(y,\varepsilon) \cap \partial\hat{H}_{r_{n_k}} = \emptyset$ eventually with positive probability.

Notice now that the events $I_k = \{X_k \in \mathcal{B}(y,\varepsilon)\}$, $\mathbb{P}(I_k) > 0$ are mutually independent. Therefore the Borel-Cantelli Lemma implies the existence, with probability 1, of a subsequence $X_{n_k} \in \mathcal{B}(y,\varepsilon)$ so that $\mathcal{B}(y,\varepsilon) \cap \hat{H}_{r_{n_k}} \neq \emptyset$. But $\mathcal{B}(y,\varepsilon)$ is a connected set, so that $y \in \partial S$ is such that $\mathcal{B}(y,\varepsilon) \subset \mathring{\hat{H}}_{r_{n_k}}$, with probability 1 for k large enough. This is impossible since $\mathring{\hat{H}}_{r_{n_k}} \subset S \oplus r_{n_k}\mathcal{B}$ and $r_{n_k} \to 0$ *a.s.* Thus (17) is proved.

Finally, in view of (13), we have

$$\partial\hat{H}_{r_n} \subset \partial S \oplus r_n\mathcal{B} \text{ e.a.s.}, \tag{18}$$

which, combined with (17), proves (5).

3.2 Proof of Theorem 2

From (6), the assumptions of Theorem 3 in [12] are fulfilled by the sequence of radiuses (r_n), then we have $S \subset \hat{S}_{r_n/4}$ *e.a.s.* Hence, Theorem 1 applies, i.e. (4),(5) and the inclusion (18) hold true.

Let then $x \in \partial S$; obviously two cases can occur. First, according to Theorem 3 in [12], *e.a.s.*, for all $x \in \partial S \cap \hat{H}^c_{r_n}$, there exists an observation $X_i \in \mathcal{X}_n$ such that

$$\|X_i - x\| \leq \left(\frac{2\ln(n)}{\delta\omega_d n}\right)^{1/d},$$

which, in view of (6), implies $d(x,\hat{H}_{r_n}) \leq r_n \leq C_S\, r_n$, since $C_S \geq 1$ (from Definition 6), thus (7) holds true. Secondly, if $x \in \partial S \cap \hat{H}_{r_n}$, we proceed by contradiction and suppose that $d(x,\partial\hat{H}_{r_n}) > C_S\, r_n$. Due to (18), this implies:

$$\mathcal{B}(x, C_S\, r_n) \subset \overbrace{S \oplus r_n\mathcal{B}}^{\circ}.$$

Therefore we have $d(x, \partial(S \oplus r_n\mathcal{B}) > C_S r_n$. Since, $x \in \partial S$ this is in contradiction with Definition 6. This concludes the proof of Theorem 2.

4 Proof of Theorem 3

4.1 Technical lemmas

Lemma 1. *Under the assumptions of Theorem 3 on S and $\mathbb{P}_X$, we have*

$$\limsup_{n\to\infty} \left(\frac{n}{\ln n}\right)^{1/(d+\alpha)} d(\mathcal{X}_n, S) \leq \left(\frac{4}{C_f\,\omega_d}\right)^{1/d+\alpha} \quad a.s., \tag{19}$$

Proof. Let X be a random variable of distribution $\mathbb{P}_X$, with density f.

Let $x \in S$. We are going to bound the probability $\mathbb{P}_X(\mathcal{B}(x, r_n))$ for any sequence $r_n \to 0$. To this aim, let us denote $t_0 = d(x, \partial S)$, and $x^* \in \partial S$ such that $\|x - x^*\| = d(x, \partial S)$. For $t \in [0, t_0 + r_n]$ let us denote $z(t) = x^* + t\,u_{x^*}$, $\mathcal{A}(t) = \mathcal{S}(O_y^+, R_S) \cap \mathcal{B}(z(t), r_n)$ and $A(t)$ its $(d-1)$-dimensional measure. We have

$$\mathbb{P}_X(\mathcal{B}(x, r_n)) \geq \int_{\max(0, t_0 - r)}^{t_0 + r_n} C_f\, t^\alpha A(t)\, dt$$

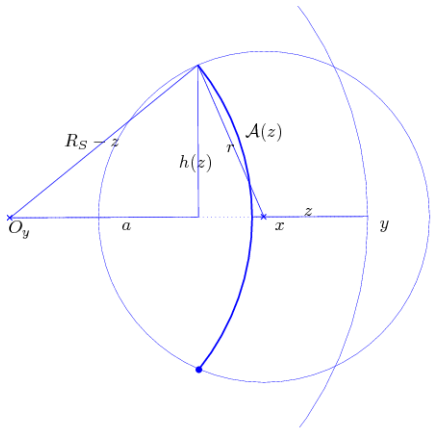


Figure 2: Probability for an observation to be in a given ball

As $r_n = o(1)$, we have $t = O(r_n)$, $t_0 = O(r_n)$ and $A(t) \geq \omega_{d-1}\, h(t)^{d-1}$ with $h(t) = t_0^2 + 2tt_0 - z^2 + r_n^2 + o(r_n^2)$, hence $h(t) = r_n^2 - (t - t_0)^2 + 2t_0^2 + o(r_n^2)$ and

$$\mathbb{P}_X(\mathcal{B}(x, r_n)) \geq \int_{t_0}^{t_0 + r_n} t^\alpha C_f\, \omega_{d-1} (r_n^2 - (t - t_0)^2 + o(r_n^2))^{\frac{d-1}{2}}\, dt$$

Using the change the variables $u = (t - t_0)^2/r_n^2$ we then have

$$\mathbb{P}_X(\mathcal{B}(x, r_n)) \geq \frac{C_f\, \omega_d}{2} r_n^{\alpha+d}(1 + o(1)). \tag{20}$$

Now, reasoning classically (as e.g. in [12]), we shall prove (19). For two given positive sequences r_n and ε_n such that $r_n \to 0$ and $\varepsilon_n \ll r_n$, let us deterministically cover S with $N(\varepsilon_n) \leq \lambda_S \varepsilon_n^{-d}$ balls of radius ε_n centered in points $x_j \in S$, $j \in \{1, \ldots, N(\varepsilon_n)\}$. We have

$$\mathbb{P}(\sup_{x\in S} \min_i ||x - X_i|| \geq r_n) \leq \mathbb{P}(\max_{x_j} \min_i ||x_j - X_i|| \geq r_n - \varepsilon_n),$$

and, noticing that $\sup_{x\in S} \min_i ||x - X_i|| = d_H(S, \mathcal{X}_n)$, and making use of (20), we obtain:

$$\mathbb{P}(d_H(S, \mathcal{X}_n) \geq r_n) \leq \lambda_S \varepsilon_n^{-d} \left(1 - \frac{C_f\, \omega_d}{2}(r_n - \varepsilon_n)^{d+\alpha}(1 + o(1))\right)^n.$$

Setting

$$r_n(u) = \left(\frac{2(\frac{d}{d+\alpha} + 1 + u)\ln n}{C_f\, \omega_d\, n}\right)^{1/(d+\alpha)}, \quad \varepsilon_n = \left(\frac{1}{n}\right)^{1/(d+\alpha)},$$

we finally have

$$\mathbb{P}(d_H(\mathcal{X}_n, S) \geq r_n(u)) \lessapprox n^{-1-u}.$$

Applying the Borel-Cantelli lemma and noticing that $\frac{d}{d+\alpha} + 1 \leq 2$ concludes the proof. □

The next lemma gives a bound on the tangent cylinders to the support of the distribution that do not contain any observations of the sample $\mathcal{X}_n$.

Lemma 2. *Assume the hypotheses of Theorem 3. Then, for any $\lambda > 0$ and $\mu > 1$ such that*

$$c = \frac{\lambda\, C_f\, \omega_{d-1}\, (\mu - 1)^{\alpha+1}}{(\alpha + 1)\, R_S^{\alpha+1}} - \frac{2d - 2}{d + 1 + 2\alpha} > 1,$$

let us denote t_n and h_n two sequences such that

$$t_n = \left(\lambda \frac{\ln n}{n}\right)^{\frac{1}{d+1+2\alpha}} (1 + o(1)), \qquad h_n = \mu \frac{r_n^2}{2R_S}(1 + o(1)).$$

Then, e.a.s, we have: for all $x \in \partial S$, $\mathcal{C}(x, t_n, h_n) \cap \mathcal{X}_n \neq \emptyset$.

Proof. First notice that, if X is drawn from a distribution $\mathbb{P}_X$ with density f, then for all $x \in \partial S$, for all $r > 0$, for all $h \geq R_S - \sqrt{R_S^2 - r^2}$ we have:

$$\mathbb{P}_X(\mathcal{C}(x, r, h)) \geq C_f \, \omega_{d-1} \int_0^{r_0} z^\alpha \left(1 - \frac{z}{R_S}\right)^{d-1} dz, \tag{21}$$

where

$$r_0 = R_S \, \frac{h - R_S + \sqrt{R_S^2 - r^2}}{\sqrt{R_S^2 - r^2}}.$$

The calculation is left to the reader as it is similar to the one used in previous Lemma.

Now, let us denote $\varepsilon_n = (\ln n)^{-2}$ and let us cover ∂S with $N(\partial S, \varepsilon_n r_n^2)$ small deterministic balls, centered in points $x_i \in \partial S$ and that have a radius $\varepsilon_n r_n^2$.

If there exists $x \in \partial S$ such that $\mathcal{C}(x, r_n, h_n) \cap \mathcal{X}_n = \emptyset$ then exists a x_i such that $x \in \mathcal{B}(x_i, \varepsilon_n r_n^2)$. and $||u_x - u_{x_i}|| \leq R_S^{-1} \varepsilon_n r_n^2$ (according to Walther 99 Th1). Thus, according to Property 8 we can find explicit values for r'_n and h'_n such that: $\mathcal{C}(x_i, r'_n, h'_n) \subset \mathcal{C}(x, r_n, h_n)$, $r'_n = r_n(1+o(1))$ and $h'_n = h_n(1+o(1))$, hence we have

$$p_n = \mathbb{P}_X(\exists x \in \partial S, \mathcal{C}(x, r_n, h_n) \cap \mathcal{X}_n = \emptyset) \leq \mathbb{P}_X(\exists x_i, \mathcal{C}(x_i, r'_n, h'_n) \cap \mathcal{X}_n = \emptyset).$$

According to (21), replacing h_n and r_n by there given values we obtain:

$$p_n \leq N(\partial S, \varepsilon_n r_n^2) \left(1 - \frac{C_f \, \omega_{d-1} \, (\mu - 1)^{\alpha+1} \, \lambda}{(\alpha + 1) \, (2R_S)^{\alpha+1}} \, \frac{\ln n}{n} \, (1 + o(1))\right)^n,$$

i.e.

$$p_n \leq \lambda_{\partial S} \varepsilon_n^{-d+1} \, (\ln n)^{-2d+2} \, \lambda^{-2\frac{d-1}{d+1+2\alpha}} \, n^{\frac{2d-2}{d+1+2\alpha}} \, n^{-\frac{C_f \omega_{d-1} (\mu-1)^{\alpha+1} \lambda}{(\alpha+1)(2R_S)^{\alpha+1}} + o(1)}.$$

With the chosen value for ε_n, we have

$$p_n \leq \lambda_{\partial S} \lambda^{-\frac{2d-2}{d+1+2\alpha}} \, n^{-c+o(1)}.$$

Since $c > 1$ we have $\sum p_n < \infty$ and we can apply the Borrel-Cantelli Lemma to conclude. □

4.2 Proof of Theorem 3

In view of the decomposition (13), we will proceed through two steps. First, let $x \in \hat{H}_{r_n} \cap S^c$, and, in view of Proposition 7, let $x^* \in \partial S$ such that

$\|x - x^*\| = d(x, \partial S)$. From Proposition 7, x, x^*, $O^+_{x^*}$ and $O^-_{x^*}$ are on the same line. Moreover, since $x \in \hat{H}_{r_n}$, the definition (2) implies the existence of $i \in \{1, \dots, n\}$ and $k+1 \leq d+1$ indexes $\{i_1, \dots, i_{k+1}\}$ such that $X_{i_j} \in \mathcal{B}(X_i, r_n)$, $j = 1 \dots k+1$, and $x \in \mathcal{H}(\{X_{i_1}, \dots, X_{i_{k+1}}\})$. Therefore, for all $j \in \{i_1, \dots, i_{k+1}\}$, we have $\|x - X_{i_j}\| \leq 2r_n$, hence $x \in \mathcal{H}(\mathcal{B}(x, 2r) \cap \mathcal{B}^c(O^-_{x^*}, R_S))$. Then, as described in Figure 3,

$$\forall x \in \hat{H}_{r_n} \cap S^c,\ d(x, \partial S) \leq R_S - \sqrt{R_S^2 - 4r_n^2}.$$

Let us assume, with no loss of generality due to the asymptotic nature of our result, that n is large enough in order to have $r_n < R_S/4$. This inclusion and (8) yield

$$\hat{H}_{r_n} \cap S^c \subset \partial S \oplus \left(\frac{2r_n^2}{R_S}(1 + o(1)) \right) \mathcal{B}. \tag{22}$$

Notice that proof of (22) is purely geometric and deterministic.

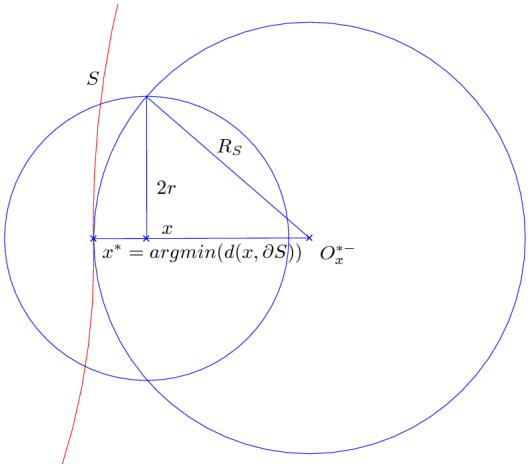


Figure 3: $x \in \mathcal{H}(\mathcal{B}(x, 2r_n) \cap \mathcal{B}^c(O^-_x, R_S)) \Longrightarrow d(x, \partial S) \leq R_S - \sqrt{R_S^2 - 4r_n^2}$

We now move to the study of $\hat{H}^c_{r_n} \cap S$. We are going to prove the following:

$$\hat{H}^c_{r_n} \cap S \subset \partial S \oplus b\, r_n^2 \mathcal{B} \ \ e.a.s., \tag{23}$$

where the constant b is explicitly given by

$$b = \frac{19}{32 R_S} + \frac{1}{2} \left(\frac{(3d - 1 + 2\alpha)(\alpha + 1) 4^{d+1+2\alpha}}{(d + 1 + 2\alpha) \lambda C_f \omega_d} \right)^{1/(\alpha+1)}. \tag{24}$$

To this aim, we argue by contradiction, and suppose that $\hat{H}^c_{r_n} \cap S \nsubseteq \partial S \oplus b r_n^2 \mathcal{B}$, i.e. there exists a point $x \in \hat{H}^c_{r_n} \cap S$ such that $d_n = d(x, \partial S) > b r_n^2$.

Let then x^* be the point of ∂S such that $d(x, x^*) = d_n$. Proceeding as in the proof of Theorem 1, we obtain that, since $x \in \hat{H}_{r_n}^c$, there exists a unit vector u such that for all X_i in $\mathcal{B}(x, r_n/2) \cap \mathcal{X}_n$, $\langle u, X_i - x \rangle \geq 0$. Let us denote

$$\rho_n = \left(\frac{4 \ln n}{C_f \omega_d n} \right)^{\frac{1}{d+\alpha}}.$$

Note that $\rho_n \ll r_n$. Three cases can then occur, which will be studied separately :

i) $d_n > 3\rho_n$: we have $\mathcal{B}(x, d_n) \subset S$ so that $\mathcal{B}(x, 3\rho_n) \subset S$. Setting $z = x - (3\rho_n/2)u$, we have (u being a unit vector) $\|x - z\| = 3\rho_n/2$, hence $z \in S$. For n large enough so that $3\rho_n \leq r_n/2$, we thus have $\mathcal{B}(z, 3\rho_n/2) \cap \mathcal{X}_n = \emptyset$. This is not possible *e.a.s.* according to Lemma 1.

ii) $d_n \leq 3\rho_n$ and $u = u_{x^*}$: this case is represented in Figure 4. We will build a tangent cylinder to ∂S which does not intersect $\mathcal{X}_n$. As described by Figure 4, we have:

$$\mathcal{C}(x^*, \sqrt{r_n^2/4 - 4d_n^2}, d_n) \cap \mathcal{X}_n = \emptyset. \tag{25}$$

Let $t_n = \sqrt{r_n^2/4 - 4d_n^2}$. Since $d_n \leq 3\rho_n$, we can write $t_n = (r_n/2)(1 + o(1)), b\, r_n^2)$. Then, in view of (25) and the fact that $d_n > b\, r_n^2$, we have

$$\mathcal{C}(x^*, t_n, b\, r_n^2) \cap \mathcal{X}_n = \emptyset,$$

which, taking (24) into account, is *e.a.s.* impossible due to Lemma 2.

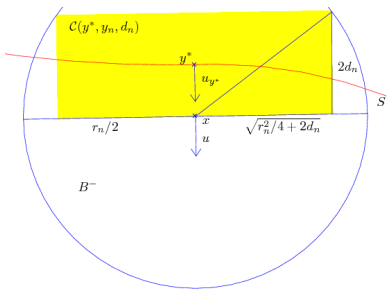


Figure 4: $X_i \in \mathcal{B}(x, r_n/2) \Longrightarrow \langle u, X_i - x \rangle \geq 0$, ie $X_i \in B^-$.

iii) $d_n \leq 3\rho_n$ and $u \neq u_{x^*}$: let $w = -u + \langle u, u_{x^*} \rangle u_{x^*}$, $v = w/||w||$ and $z^* = x^* + (\sqrt{r_n^2/4 - 4d^2}/2)v$. As displayed in Figure 5, we have

$$\mathcal{C}^{u_{x^*}}(z^*, \sqrt{r_n^2/4 - 4d_n^2}/2, d_n) \cap \mathcal{X}_n = \emptyset.$$

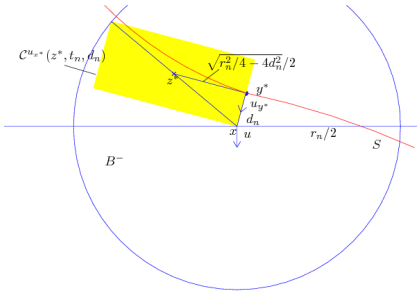


Figure 5: $X_i \in \mathcal{B}(x, r_n/2) \Longrightarrow \langle u, X_i - x\rangle \geq 0$, ie $X_i \in B^-$.

Let $t_n = \sqrt{r_n^2/4 - 4d_n^2}/2 = (r_n/4)(1 + o(1))$. As in the previous case, we deduce that

$$\mathcal{C}^{u_{x^*}}(z^*, t_n, br_n^2) \cap \mathcal{X}_n = \emptyset. \tag{26}$$

The cylinder featured in (26) is not a tangent one, then the proof is yet to be completed. From Definition 7 (see Figure 6), we have $d(z^*, \partial S) \leq \alpha_n = -R_S + \sqrt{R_S^2 + r_n^2/16} \sim r_n^2/(32R_S)$.

For a given $z \in \partial S$ such that $d(z, z^*) \leq \alpha_n$, we have $d(z, x^*) \leq \alpha_n + r_n/4$, so that, according to Proposition 5, $||u_z - u_{x^*}|| \leq (r_n/4 + \alpha_n)/R_S$. Let

$$\begin{aligned} e_1 &= \frac{r_n^2}{32R_S} + 2\left(\sqrt{t_n^2 + b^2 r_n^4} + \frac{r_n^2}{32R_S}\right)\frac{1}{R_S}\left(\alpha_n + \frac{r_n}{4}\right), \\ e_2 &= \frac{r_n^2}{32R_S} + \sqrt{t_n^2 + b^2 r_n^4}\,\frac{1}{R_S}\left(\alpha_n + \frac{r_n}{4}\right). \end{aligned}$$

Proposition 8 yields:

$$\mathcal{C}(z, t_n - e_1, br_n^2 - e_2) = \mathcal{C}^{u_z}(z, t_n - e_1, br_n^2 - e_2) \subset \mathcal{C}^{u_{x^*}}(z^*, t_n, br_n^2).$$

A basic calculation gives

$$t_n - e_1 = t_n(1 + o(1)), \quad br_n^2 - e_2 = \left(b - \frac{3}{32R_S}\right) r_n^2\,(1 + o(1)),$$

and, in view of (26), we have

$$\mathcal{C}(z, t_n - e_1, br_n^2 - e_2) \cap \mathcal{X}_n = \emptyset,$$

which, arguing as in the previous case, is *e.a.s.* impossible. Therefore (23) is proved.

Notice now that, from Proposition 6, S is partly expandable with expandability constant $C_S = 1$. Taking this into account, and reasoning as in the proof of Theorem 2, we obtain (9) and (10) as a direct consequence of inclusions (22) and (23).

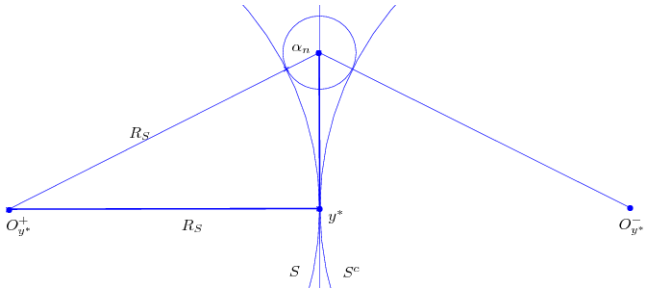


Figure 6: $d(z^*, \partial S) \leq \alpha_n$.

5 Proof of Theorem 4

5.1 Technical framework and lemmas

We will first introduce some concepts related to the geometrical nature of the estimator $\hat{H}_{r_n}$.

Definition 11. *A $d'-$dimensional convex polyhedron in $\mathbb{R}^d$ is the convex hull of a set of $k > d'+1$ points $\{x_1, \dots x_k\}$ which spans $\mathbb{R}^{d'}$.*

Definition 12. *A $d'-$dimensional polyhedron is a finite union of $d'-$dimensional convex polyhedrons. Its boundary ∂A is a $(d'-1)-$dimensional polyhedron. There exist $N_\partial \in \mathbb{N}$ and a set $\{\sigma_1, \dots, \sigma_{N_\partial}\}$ of $(d'-1)-$dimensional convex polyhedrons, called faces, such that $\partial A = \bigcup \sigma_i$. Let $\mathring{\sigma}_i = \{x \in \sigma_i, j \neq i \Rightarrow x \notin \sigma_j\}$ be the interior of the face for the usual topology induced by σ_i, and v_i the unit normal vector to σ_i pointing outward A. Then, for all $z \in \mathring{\sigma}_i$ there exist $\varepsilon_z > 0$ such that :*

i) $\forall \varepsilon \in]0, \varepsilon_z]$, $z + \varepsilon v_i \notin A$,

ii) $\forall \varepsilon \in]0, \varepsilon_z]$, $z - \varepsilon v_i \in \mathring{A}$.

Lemma 3. *Let A be a $d-$dimensional polyhedron in $\mathbb{R}^d$. Let $z \in \partial A$ and $u \in \mathbb{R}^d$.*

i) If there exists $\varepsilon_0 > 0$ such that $z + \varepsilon u \in A^c$ for all $\varepsilon \in]0, \varepsilon_0]$ then, there exists $i \in \{1, \dots N_\partial\}$ such that $z \in \sigma_i$ and $\langle v_i \, , \, u \rangle \geq 0$;

ii) if there exists $\varepsilon_0 > 0$ such that $z + \varepsilon u \in \mathring{A}$ for all $\varepsilon \in]0, \varepsilon_0]$ then, there exists $i \in \{1, \dots N_\partial\}$ such that $z \in \sigma_i$ and $\langle v_i \, , \, u \rangle \leq 0$.

Proof. Let us first notice that the result is obvious when the point z belongs to a unique face of ∂A, that is $z \in \mathring{\sigma}_i$ for a unique index $i \in \{1, \ldots, N_\partial\}$.

It then remains to prove the result for the points z located at the intersection of more than one face. Let z be a point of ∂A and $I \subset \{1, \ldots, N_\partial\}$ the set of indexes such that $z \in \bigcap_{i\in I} \sigma_i,\ z \notin \bigcup_{i\in I^c} \sigma_i$. Let us introduce $a = \min_{i\in I^c} d(z, \sigma_i)$. We will only prove the first assertion of the lemma (the proof of the second one being similar). Assume that there exists $\varepsilon_0 > 0$ such that $z + \varepsilon u \in A^c$ for all $\varepsilon \in]0, \varepsilon_0]$ and let us suppose that there also exists $i_0 \in I$ such that $\langle u, v_{i_0}\rangle < 0$. We set $\varepsilon = \min(\varepsilon_0, a/4)$ and $z_1 = z + \varepsilon u \in A^c$. For all $z_2 \in \mathring{\sigma}_{i_0}$ with $||z - z_2|| < \varepsilon$, we have

$$\langle z_2 - z_1, v_{i_0}\rangle = \varepsilon\langle u, v_{i_0}\rangle < 0. \tag{27}$$

According to our remark at the beginning of the proof, one can find $t(z_2) \in]z_1, z_2[\cap\mathring{A}$. Thus, there exists $x(z_2) \in]z_1, t(z_2)[\cap\partial A$ such that

$$[t(z_2), x(z_2)[\subset \mathring{A}. \tag{28}$$

Since $||x(z_2) - z|| \leq ||x(z_2) - z_2|| + ||z_2 - z_1|| \leq 3\varepsilon < a$, there exists $j_0 \in I$ such that $x(z_2) \in \sigma_{j_0}$. Noticing that $x(z_2) = \sigma_{j_0} \cap [z_1, z_2]$, we can choose z_2 such that $x(z_2) \in \mathring{\sigma}_{j_0}$. Hence, (27), (28) and our initial remark allow to conclude that choosing $i = j_0$ proves the first assertion of the Lemma. □

Lemma 4. *Assume the hypotheses of Theorem 3 on $\mathcal{X}_n$, $\mathbb{P}_X$ and S. If the sequence of radiuses r_n satisfies* (8)*, then there exist, e.a.s., at least $d + 1$ observations in every $\mathcal{B}(X_i, r_n)$, $i = 1 \ldots n$.*

Proof. For a given $i \in \{1, \ldots, n\}$, let us denote $p_n(i) = \mathbb{P}_X(\mathcal{B}(X_i, r_n))$. The probability $q_n(i)$ for $\mathcal{B}(X_i, r_n)$ to contain less than d other observations of the sample is then given by

$$q_n(i) = \sum_{k=0}^{d-1} \binom{n}{k} p_n(i)^k (1 - p_n(i))^{n-k}.$$

For $n > 2d$, we have

$$q_n(i) \leq \frac{n^{d-1}}{(d-1)!} (1 - p_n(i))^n \sum_{k=0}^{d-1} \left(\frac{p_n(i)}{1 - p_n(i)}\right)^k,$$

then

$$q_n(i) \leq \frac{n^{d-1}}{(d-1)!} (1 - p_n(i))^n \frac{1 - p_n(i)}{1 - 2p_n(i)}.$$

As in the proof of Lemma 1, we have $p_n(i) \geq A_0 r_n^{d+\alpha}$. Moreover, it is obvious that $p_n(i) \leq \omega_d \max(f) r_n^d \to 0$, hence, setting $u = \frac{\alpha+1}{d+1+2\alpha}$, and in view of (8), we obtain

$$q_n(i) \lessapprox n^{d-1} \exp\left(-A_0 \lambda^{d+\alpha} n^u (\ln n)^{1-u}\right).$$

Finally, the probability P_n that theres exists $i \in \{1, \ldots, n\}$ such that $\mathcal{B}(X_i, r_n)$ contains less than $d+1$ observations of the sample satisfies:

$$P_n \lessapprox n^d \exp\left(-A_0 \lambda^{d+\alpha} n^u (\ln n)^{1-u}\right).$$

Obviously, we have $\sum P_n < \infty$ and the application of the Borrel-Cantelli lemma concludes the proof. □

Corollary 1. *Assume the hypotheses of Theorem 3 on $\mathcal{X}_n$, $\mathbb{P}_X$ and S. Then, if the sequence of radiuses r_n satisfies (8), the estimator $\hat{H}_{r_n}$ is e.a.s. a $d-$dimensional manifold and a polyhedron.*

The proof is a consequence of the fact than $d+1$ points drawn from a probability measure which is absolutely continuous with respect to the Lebesgue measure necessarily span $\mathbb{R}^d$ a.s.

Lemma 5. *Assume the hypotheses of Theorem 3 on $\mathcal{X}_n$, $\mathbb{P}_X$ and S. Assume that the sequence of radiuses r_n satisfies (8). There exists e.a.s. a finite number $\sigma_1, \ldots, \sigma_{N_\partial}$ of $(d-1)-$dimensional polyhedrons such that:*

i) $\partial \hat{H}_{r_n} = \bigcup_{i=1}^{N_\partial} \sigma_i$,

ii) for all $i \in \{1, \ldots, N_\partial\}$ there exists a set of d observations $\{X_{1,i}, \ldots, X_{d,i}\} \subset \mathcal{X}_n$ and a unique unit vector v_i, normal to σ_i and outward to $\hat{H}_{r_n}$, such that

ii-a) $\sigma_i \subset \mathcal{H}(\{X_{1,i}, \ldots, X_{d,i}\})$,

ii-b) there exists i^* such that $\{X_{1,i}, \ldots, X_{d,i}\} \subset \mathcal{B}(X_{i^*}, r_n)$,

ii-c) for all j such that $\{X_{1,i}, \ldots, X_{d,i}\} \subset \mathcal{B}(X_j, r_n)$, for all k such that $X_k \in \mathcal{B}(X_j, r_n)$, for all $x \in \sigma_i$, we have $\langle x - X_k, v_i \rangle \geq 0$.

Proof. Point i) is a direct consequence of Corollary 1.

For all $x \in \partial \hat{H}_{r_n}$, there exists an index i_x such that $x \in \partial \mathcal{H}(\mathcal{B}(X_{i_x}, r_n) \cap \mathcal{X}_n)$ and therefore there exist $A_i := \{X_{1,i_x}, \ldots, X_{d,i_x}\} \subset \mathcal{B}(X_{i_x}, r_n) \cap \mathcal{X}_n$, such that $x \in \mathcal{H}(A_i)$.

Let $i \in \{1, \ldots, N_\partial\}$ such that $x \in \sigma_i$. Necessarily, $\sigma_i \subset \mathcal{H}(A_i)$; otherwise there would be more than d observations belonging to a $(d-1)-$dimensional plane, which is *a.s.* impossible. This proves point $ii-a$).

Choosing $i^* = i_x$ then proves point $ii - b)$.

Now, Let us define v_i as the unit vector, orthogonal to the plane which contains the A_i, and pointing outward $\mathcal{H}(\mathcal{B}(X_{i_x}, r_n) \cap \mathcal{X}_n)$. Then, for all j such that $X_j \in \mathcal{B}(X_{i_x}, r_n) \cap \mathcal{X}_n$, we obviously have $\langle y - X_j, v_i \rangle \geq 0, \quad \forall y \in \sigma_i$. Point ii-c) than holds in a same manner, σ_i being a face of $\partial \hat{H}_{r_n}$. □

Lemma 6. *Assume the hypotheses of Theorem 3 on $\mathcal{X}_n$, $\mathbb{P}_X$ and S. Assume that the sequence of radiuses r_n satisfies (8). Then, e.a.s., for all $x \in \partial \hat{H}_{r_n}$, for all $i \in \{1, \ldots, N_\partial\}$ such that $x \in \sigma_i$, for all $z \in \partial S$ satisfying $\|x - z\| = d(x, \partial S)$, we have $\langle v_i, u_z \rangle < 0$.*

Proof. Let $\rho_0 > 0$ be large enough. Let

$$\begin{aligned} \rho_n &= 2\rho_0 \left(\frac{\ln n}{n} \right)^{\frac{1}{d+\alpha}}, \\ \varepsilon_n &= \varepsilon_0 \left(\frac{\ln n}{n} \right)^{\frac{2}{d+1+2\alpha}}, \end{aligned}$$

where $\rho_0 > 0$ is large enough so that, according to Lemma 1, *e.a.s.*, for any $x \in S$, we have $\mathcal{B}(x, \rho_n) \cap \mathcal{X}_n \neq \emptyset$. The number $\varepsilon_0 > 0$ is chosen large enough so that, in view of (10),

$$d_H(\partial \hat{H}_{r_n}, \partial S) \leq \varepsilon_n, \quad e.a.s. \tag{29}$$

Let us proceed by contradiction. Suppose that there exists $i \in \{1, \ldots, N_\partial\}$ and $x \in \sigma_i$ such that $\langle v_i, u_z \rangle \geq 0$. According to Lemma 5 there exists X_{i^*} and $\{X_{1,i}, \ldots, X_{d,i}\} \subset \mathcal{B}(X_{i^*}) \cap \mathcal{X}_n$ such that $\sigma_i \subset \mathcal{H}(\{X_{1,i}, \ldots, X_{d,i}\})$.

In a first step, let us prove that we *e.a.s.* have

$$\langle v_i, u_z \rangle \geq 0 \Longrightarrow \langle X_{i^*} - x, v_i \rangle > r_n - 2\rho_n. \tag{30}$$

Once again we proceed by contradiction and suppose that $\langle X_{i^*} - x, v_i \rangle \leq r_n - 2\rho_n$. Let $y = X_{i^*} + (r_n - \rho_n)v_i$; we have $\mathcal{B}(y, \rho_n) \cap \mathcal{X}_n = \emptyset$. In view of Definition 7, the triangle inequality and (29), we have,

$$X_{i^*} \in \mathcal{B}(z, r_n + \varepsilon_n) \cap \mathcal{B}^c(O_z^-, R_S), \quad e.a.s.$$

That implies

$$X_{i^*} \in \mathcal{B}(z, r_n + \varepsilon_n) \cap \{t, \langle t - z, u_z \rangle \geq -e_n\}, \quad e_n = \frac{r_n^2}{2R_s}(1 + o(1)).$$

Noticing that $||X_{i^*} - y|| \leq r_n$ and $\langle y - X_{i^*}, u_z \rangle = (r_n - \rho_n)\langle v_i, u_z \rangle \geq 0$, we obtain:

$$y \in \mathcal{B}(z, 2r_n + \varepsilon_n) \cap \{t, \langle t - z, u_z \rangle \geq -e_n\}.$$

Therefore $||O_z^+ - y|| \leq R_S + e'_n$ with $e'_n \sim \frac{5r_n^2}{2R_S}$ (see Figure 7). Let now

$$\tilde{y} = \begin{cases} y \text{ if } y \in S \\ \text{argmin}\{w \in \partial S, d(w, y)\} \text{ if } y \notin S \end{cases}$$

The previous considerations imply that $||\tilde{y} - y|| \leq e'_n$, that is there exists $\tilde{y} \in S$ such that $\mathcal{B}(\tilde{y}, \rho_n - e'_n) \cap \mathcal{X}_n = \emptyset$ which is, *e.a.s.*, impossible. That proves (30).

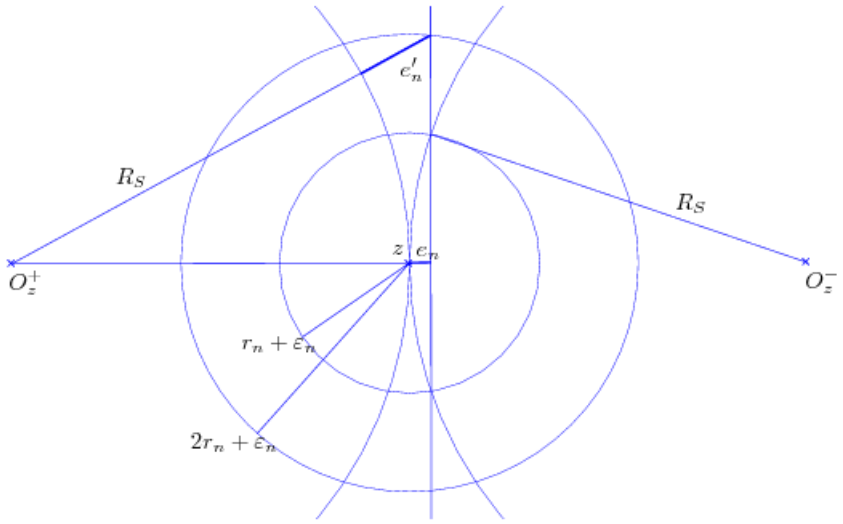


Figure 7: $X_{i^*} \in \mathcal{B}(z, r_n + \varepsilon_n) \cap \mathcal{B}^c(O_z^-, R_S) \Rightarrow d(y, O^+_{\varphi_{r_n}(x)}) \leq R_S + e'_n)$

But now (30) implies that, for any k and l, $||X_{k,i} - X_{l,i}|| \leq 4\sqrt{\rho_n r_n - r_n^2}$ (see Figure 8). Therefore, for all k, $||x - X_{k,i}|| \leq 4\sqrt{\rho_n r_n - r_n^2}$. Finally, when n is large enough we have: $\{X_{1,i}, \ldots, X_{d,i}\} \subset \mathcal{B}(X_{1,i}, r_n)$ and $||x^* - X_{1,i}|| \leq 4\sqrt{\rho_n r_n - r_n^2} \ll r_n - 2\rho_n$ which is *e.a.s.* not possible according to (30). □

5.2 Proof of Theorem 4

Let us introduce the function

$$\varphi_{r_n} : \begin{cases} \partial \hat{H}_{r_n} & \longrightarrow & \partial S \\ x & \longmapsto & \text{argmin}_{y \in \partial S} \|x - y\| \end{cases}$$

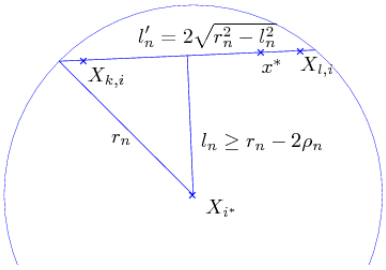


Figure 8: $\langle X_{i^*} - x^*, u_{x^*} \rangle \geq r_n - 2\rho_n \Rightarrow \forall i, j,\ ||X_{k,i} - X_{l,i}|| \leq 4\sqrt{\rho_n r_n - r_n^2}$

Proving (11) now boils down to proving that φ_{r_n} is well-defined, continuous, bijective, and that $\varphi_{r_n}^{-1}$ is continuous.

First, from Proposition 7, for any x such that $d(x, \partial S) < R_S$, there exists a unique $y \in \partial S$ that realizes $\min ||x - y||$. Moreover, the assumptions being the same as they are for Theorem 3, inclusions (22) and (23) hold, hence, for n large enough, $\partial \hat{H}_{r_n} \subset \partial S \oplus \frac{R_S}{2}\mathcal{B}$ *e.a.s.* Therefore the function φ_{r_n} is *e.a.s.* well defined.

Now, let us prove that the application

$$\varphi : \left\{ \begin{array}{ccc} \partial S \oplus \frac{R_S}{2}\mathcal{B} & \longrightarrow & \partial S \\ x & \longmapsto & \operatorname{argmin}_{y \in \partial S} \|x - y\| \end{array} \right. \tag{31}$$

is continuous, which will obviously imply the continuity of φ_{r_n}. Let $0 < \varepsilon \leq R_S/4$, and $x \in \mathbb{R}^d$ such that $d(x, \partial S) < R_S/2$. From (31), for $x' \in \mathcal{B}(x, \varepsilon)$, we have $d(x', \partial S) \leq ||x' - \varphi(x)||$, i.e. $\varphi(x') \in \mathcal{B}(x', ||\varphi(x) - x'||)$. Since balls of radius R_S roll freely inside S, we have $\mathring{\mathcal{B}}(O^+_{\varphi(x)}, R_S) \subset \mathring{S}$, which implies, $\varphi(x')$ belonging to ∂S, that $\varphi(x') \notin \mathring{\mathcal{B}}(O^+_{\varphi(x)}, R_S)$. From this we deduce, as represented on Figure 9, that $||\varphi(x') - \varphi(x)|| \leq 4\varepsilon$. This proves the continuity of φ.

From the inclusions (22) and (23), there exists a constant a such that $\hat{H}_{r_n} \Delta S \subset \partial S \oplus a r_n^2 \mathcal{B}$ *e.a.s.* Let $x \in \partial S \cap \hat{H}^c_{r_n}$ (respectively $x \in \partial S \cap \hat{H}_{r_n}$), and $y = x + a r_n^2 u_x$ (resp. $y = x - a r_n^2 u_x$). We have $y \in \hat{H}_{r_n}$ (resp. $y \in \hat{H}^c_{r_n}$) *e.a.s.*, hence according to Propositions 2 and 7, $[x, y]$ intersects $\partial \hat{H}_{r_n}$ at a point x^* that satisfies $\varphi_{r_n}(x^*) = x$. Therefore φ_{r_n} is surjective.

Let us now prove that φ_{r_n} is injective. Arguing by contradiction, let us suppose that there exists two points, x and y in $\partial \hat{H}_{r_n}$ such that $\varphi_{r_n}(x) = \varphi_{r_n}(y) = z$. Then, from Proposition 7, x, y and z belong to the same line directed by u_z.

Let us now consider different cases. Let us recall that, in view of Definition 12, a point in $\partial \hat{H}_{r_n}$ belongs to a face σ_i with outward unit normal vector v_i.

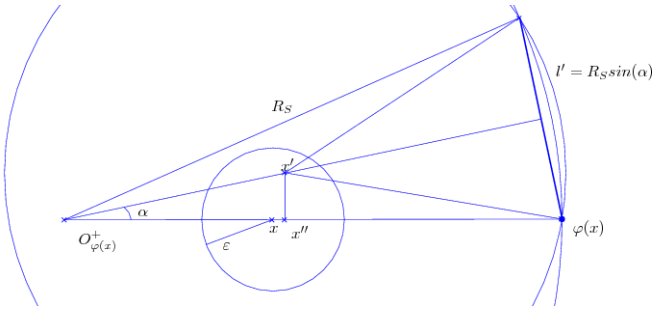


Figure 9: $||x - O^+_{\varphi(x)}|| \geq R_S/2$ and $\varepsilon \leq R_S/4 \Rightarrow \sin(\alpha) = \frac{||x'-x''||}{||O^+_{\varphi(x)}-x'||} \leq \frac{4\varepsilon}{R_S} \Rightarrow l' \leq 4\varepsilon$

i) The first one is when $[x, y] \cap \partial\hat{H}_{r_n} = [x, y]$. In this case, consider the point $x^* = (x+y)/2$. Then, x^* belongs to a face $\sigma_i \subset \partial\hat{H}_{r_n}$ that contains the vector u_z. This last point implies that $\langle v_i\,,\, u_z\rangle = 0$, which is, *e.a.s.*, impossible according to Lemma 6.

ii) The second case is when $[x, y] \cap \partial\hat{H}_{r_n} \neq [x, y]$ and $[x, y] \cap \hat{H}^c_{r_n} \neq \emptyset$. There exists $z^* \in]x, y[\cap\hat{H}^c_{r_n}$. Let us define the closed half line $\Delta^* = \{z^* - \lambda u_z, \lambda \geq 0\}$ and the point $\tilde{z} = \text{argmin}\{||t - z^*||, t \in \hat{H}_{r_n} \cap \Delta^*\}$. Recall that $u_{\tilde{z}} = u_z$. Note that $\tilde{z} \in \partial\hat{H}_{r_n}$, and that, for $0 < \varepsilon < ||\tilde{z} - z||$, $\tilde{z} + \varepsilon u_z \in \hat{H}^c_{r_n}$. Thus, in view of Lemma 3, there exists a face $\sigma_i \subset \partial\hat{H}_{r_n}$ (with normal vector v_i) which contains $\tilde{z}$ such that $\langle v_i\,,\, u_{\tilde{z}}\rangle = \langle v_i\,,\, u_z\rangle \geq 0$. This is impossible according to Lemma 6.

iii) The third case that is $[x, y] \cap \partial\hat{H}_{r_n} \neq [x, y]$ and $[x, y] \cap \mathring{\hat{H}}_{r_n} \neq \emptyset$ can be solved as the previous one. Namely There exists $z^* \in]x, y[\cap\mathring{\hat{H}}_{r_n}$. Let us define the closed half line $\Delta^* = \{z^* + \lambda u_z, \lambda \geq 0\}$ and the point $\tilde{z} = \text{argmin}\{||t - z^*||, t \in \hat{H}_{r_n} \cap \Delta^*\}$. The conclusion is the same as in the previous case.

We have proved that φ_{r_n} is *e.a.s* bijective and continuous. It then remains to prove that $\varphi^{-1}_{r_n}$ is *e.a.s* continuous. For a given point $x \in \partial S$ let us denote $x^+ = x + \frac{R_S}{2}u_x$ and $x^- = x - \frac{R_S}{2}u_x$. Due to the inclusions (22) and (23), the line segment $[x^+, x^-]$ intersects $\partial\hat{H}_{r_n}$ at a unique point $\varphi^{-1}_{r_n}(x)$. Let $k \in \{1, \ldots, N_\partial\}$ be the number of faces in $\partial\hat{H}_{r_n}$ containing $\varphi^{-1}_{r_n}(x)$ (as in Definition 12). Renumbering the faces for the ease of reading, we have

$$0 = d([x^+, x^-], \sigma_1) = \ldots = d([x^+, x^-], \sigma_k),$$

and

$$0 < d([x^+, x^-], \sigma_{k+1}) \leq \ldots \leq d([x^+, x^-], \sigma_{N_\partial}).$$

From Lemma 6, for all $i \leq k$, we have $\langle -v_i, u_x \rangle > 0$ *e.a.s.* Let us then define $e_0 = \max_{i \leq k}\{1/\langle -v_i, u_x \rangle\} > 0$.

Let $\varepsilon > 0$. For $z \in \partial S \cap \mathcal{B}(x, \varepsilon)$, let $z^+ = z + \frac{R_S}{2} u_z$ and $z^- = z - \frac{R_S}{2} u_z$. Proposition 5 and the triangle inequality imply that,for all $\lambda \in [-\frac{R_S}{2}, \frac{R_S}{2}]$, $\|z_\lambda - x_\lambda\| \leq 3\varepsilon/2$ where $x_\lambda = x + \lambda u_x$ and $z_\lambda = z + \lambda u_z$. From this we deduce

$$d_H([x^+, x^-], [z^+, z^-]) \leq \frac{3}{2}\varepsilon. \tag{32}$$

Now, choose $\varepsilon < \frac{2}{3d([x^+, x^-], \sigma_{k+1})}$. For all $z \in \partial S \cap \mathcal{B}(x, \varepsilon)$, the line segment $[z^+, z^-]$ does not intersect any σ_i for $i > k$. The application φ_{r_n} being *e.a.s.* bijective, the segment $[z^+, z^-]$ intersects $\partial \hat{H}_{r_n}$ at a unique point $\varphi_{r_n}^{-1}(z) \in \sigma_i$ for some index $i \leq k$. The inequality (32) then implies that

$$\|\varphi_{r_n}^{-1}(z) - \varphi_{r_n}^{-1}(x)\| \leq \frac{3\varepsilon}{2\langle -v_i, u_x \rangle} \leq 3e_0\varepsilon/2,$$

which proves the continuity of $\varphi_{r_n}^{-1}$. Thus (11) is proved.

The proof of (12) is now straightforward. It suffices to define $\psi_{r_n} : \hat{H}_{r_n} \to S$, the natural extension of φ_{r_n} to $\hat{H}_{r_n}$, as follows:

i) if $x \in S \ominus (R_S/2)\mathcal{B}$ then $\psi_{r_n}(x) = x$,

ii) if $x \in S \setminus (S \ominus (R_S/2)\mathcal{B})$ then $\psi_{r_n}(x) = x' - \frac{2||x'-z||.||x'-x||}{R_S} u_{g(x)}$ where

   a) $g(x) = \text{argmin}(d(x, \partial S))$,

   b) $x' = g(x) - \frac{R_S}{2} u_{g(x)}$,

   c) $z = \varphi_{r_n}^{-1}(g(x))$.

It is easy (and left to the reader) to prove that ψ_{r_n} is *e.a.s.* an homeomorphism from $\hat{H}_{r_n}$ to S. This concludes the proof.

6 Numerical simulations

The aim of this section is to validate the efficiency of our method in the case of some given examples for which the support and density are prescribed.

6.1 Simulated data

First we present numerical simulations on some "toy" examples. The observations have been generated as follows.

1. 500 realizations uniformly drawn on the star shape

$$S = [-1,1]^2 \setminus (\bigcup_{i:1}^{4} \mathcal{B}(C_i, 1)),\ C_i \in \{(-1,-1),\ (-1,1),\ (1,-1),\ (1,1)\}\,.$$

   In this case S only satisfies the assumptions of Theorem 1.

2. 500 realizations drawn on an asterisk shape with the following distribution:

$$\begin{array}{l} X = (\cos(\theta)x + \sin(\theta)y, \cos(\theta)y - \sin(\theta)x), \\ \mathbb{P}(\theta = k\pi/4) = 1/4,\ k \in \{0,1,2,3\}, \\ (x,y) \hookrightarrow \mathcal{U}([-1,1] \times [-0.05, 0.05]). \end{array}$$

   In this case the assumptions of Theorem 2 are satisfied.

3. 500 realizations drawn in $S = \mathcal{B}(0,1) \setminus \mathcal{B}(0,r)$ with the following distribution:

$$X = (r\cos\theta, r\sin\theta),\ \theta \hookrightarrow \mathcal{U}([0,2\pi]),\ r \hookrightarrow \mathcal{U}([r,1]).$$

   The results for different values of r are presented ; $r = 0.9$, $r = 0.5$ and $r = 0$ (respectively denoted $3-a$), $3-b$) and $3-c$). In this case, the assumptions of Theorems 3 and 4 are fulfilled.

Figure 10 presents three series of results: the Devroye-Wise estimator for the best radius according to support estimation, the Devroye Wise estimator for the best radius according to the boundary estimation and the Local-Convex-Hull estimator for the best radius according to the support estimation.

It can be noticed that, the boundary of the Devroye-Wises estimator with the best radius according for support estimation does not accurately estimate the boundary. Also, the Devroye-Wise estimator with the best radius for boundary estimation overfills the support. On the other hand the Local Convex Hull estimator, with the best radius for support estimation provides much better estimations of the support and its boundary.

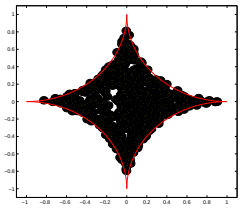
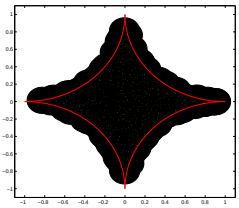
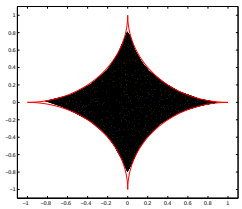
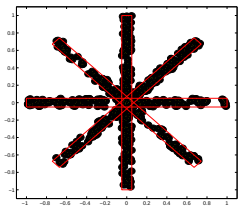
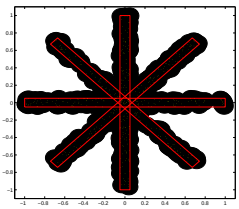
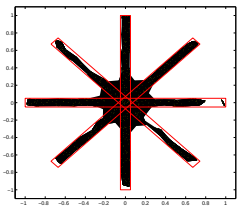
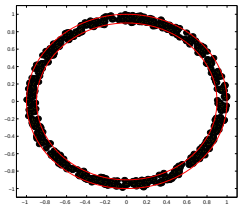
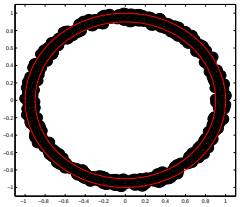
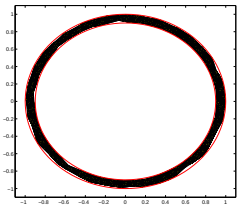
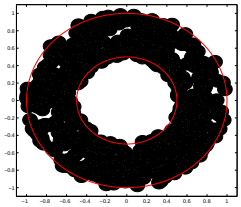
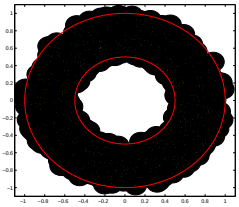
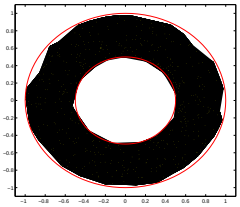
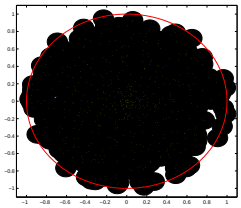
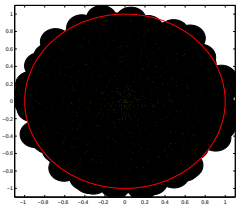
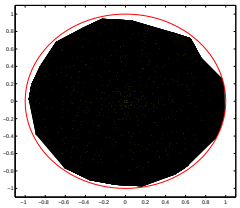
	Devroye-Wise best radius for Support.	Devroye-Wise best radius for Boundary.	Local-Convex-Hull best radius for Support.
1			
2			
$3-a)$			
$3-b)$			
$3-c)$			

Figure 10: Comparison of the estimators for the different examples

6.2 Application to a real data

Here the method is applied to a set of 5323 locations of epicenters of earthquakes with a magnitude greater than 6. In this example we have slightly

changed the estimator in order to take into account the fact that we are working with spherical coordinates. The location X_i of each epicenter is given on the sphere $\mathcal{S}^2$. On Figure 11 we represent the modified estimator:

$$\tilde{H}_r = \bigcup_i \mathcal{H}(p(\mathcal{B}(X_i, r) \cap \mathcal{X}_n)),$$

where p returns to each point of $\mathcal{S}^2$ its latitude and longitude.

In order to remove the noise from the data we also computed this estimator on the 4000 epicenters where the estimated density is the highest. The density was estimated with a nearest neighbor density estimator (see [23]).

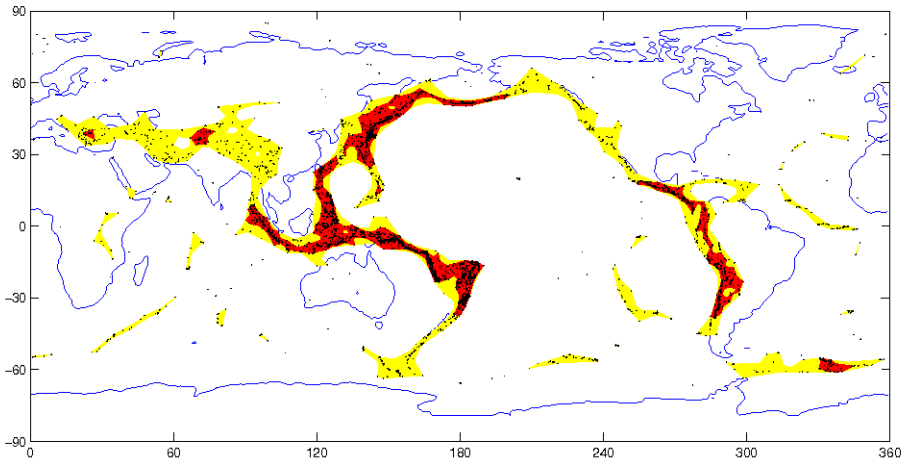


Figure 11: Results for seism epicenters : support estimation in yellow, support estimation for the 4000 epicenters of highest density in red.

References

[1] A. Baillo, A. Cuevas, and A. Justel. Set estimation and nonparametric detection. *The Canadian Journal of Statistics*, 28:765–782, 2000.

[2] I. Bárány. Random polytopes in smooth convex bodies. *Mathematika*, 39:81–92, 1982.

[3] G. Biau, B. Cadre, D.M. Mason, and B. Pelletier. Asymptotic normality in density support estimation. *Electronic Journal of Probability*, pages 2617–2635, 2009.

[4] J.D. Boissonnat and A. Ghosh. Manifold reconstruction using tangential delaunay complexes. In *In Proceedings of the 26th Annual Symposium on Computational Geometry*, 2010.

[5] P. Bubenik, G. Carlsson, P.T. Kim, and Z.M. Luo. Statistical topology via morse theory, persistence, and nonparametric estimation. contemporary mathematics, 516 (2010). *Algebraic Methods in Statistics and Probability II. Contemporary Mathematics,*, 516:75–92, 2010.

[6] G. Carlsson. Persistent homology and the analysis of high dimensional data. In *Symposium on the Geometry of Very Large Data Sets , Fields Institute for Research in Mathematical Sciences*, 2005.

[7] G. Carlsson. Topology and data. *Bulletin of the American Mathematical Society*, 46:255–308, 2009.

[8] G. Carlsson, A. Zomorodian, A. Collins, and L.J. Guibas. Persistence barcodes for shapes. *International Journal of Shape Modeling*, 11:149–187, 2005.

[9] Chevalier. Estimation du support et du contour du support d'une loi de probabilit. *Annales de l'Institut Henri Poincaré (B) Probability and Statistics*, 12:339–364, 1976.

[10] A. Cholaquidis, A. Cuevas, and R. Fraiman. On poincar cone property. *Annals of Statistics (to appear)*, 2014.

[11] A. Cuevas. On pattern analysis in the nonconvex case. *Kybernetes*, 19:26–33, 1990.

[12] A. Cuevas and A. Rodríguez-Casal. On boundary estimation. *Advanced in Applied Probability*, 36:340–354, 2004.

[13] L. Devroye and G.L. Wise. Detection of abnormal behavior via nonparametric estimation of the support. *SIAM Journal of Applied Mathematics*, 38:480–488, 1980.

[14] L. Dumbgen and G. Walther. Rates of convergence for random approximations of convex sets. *Advances in Applied Probabilitity*, 28:384–393, 1996.

[15] H. Edelsbrunner and N.R. Shah. Triangulating topological spaces. *International Journal of Computational Geometry Applications*, 7:365–378, 1997.

[16] B. Efron. The convex hull of a random set of points. *Biometrika*, 15:331–343, 1965.

[17] W.M. Getz and C.C. Wilmers. A local nearest-neighbor convex-hull construction of home ranges and utilization distributions. *Ecography*, 27:489–505, 2004.

[18] W. Hardle, B.U. Park, and A.B. Tsybakov. Estimation of non-sharp support boundary. *Journal of Multivariate Analysis*, 55:205–218, 1995.

[19] D.C. Kesler. Foraging habitat distributions affect territory size and shape in the tuamotu kingfisher. *International Journal of Zoology*, 2012, 2012.

[20] W. Khnel. *Differential Geometry: Curves - Surfaces - Manifolds*. AMS, 2005.

[21] L.M. Korte. Habitat selection at two spatial scales and diurnal activity patterns of adult female forest buffalo. *Journal of Mammalogy*, 89:115–125, 2008.

[22] Q. Liu, J. Yang, X. Yang, J. Zhao, and H. Yu. Foraging habitats and utilization distributions of black-necked cranes wintering at the napahai wetland, china. *Journal of field ornithology*, 81:21–30, 2010.

[23] D.O. Loftsgaarden and C.P. Quesenberry. A nonparametric estimate of a multivariate density function. *The Annals of Mathematical Statistics*, 36(3):1049–1051, 1965.

[24] M. Reitzner. Random polytopes and the Efron-Stein jackknife inequality,. *Annals of Probability*, 31:21362166., 2003.

[25] A. Rodríguez-Casal. Set estimation under convexity type assumptions. *Annales de l'Institut Henri Poincaré (B) Probability and Statistics*, 43:763774, 2007.

[26] R. Schneider. Random approximation of convex sets. *Journal of Microscopy*, 151:211–227, 1988.

[27] G. Walther. On a generalization of Blaschke's rolling theorem and the smoothing of surfaces. *Mathematical Methods in the Applied Sciences*, 22:301–316, 1999.

[28] A. Zomorodian and G. Carlsson. Computing persistent homology. *Discrete and Computational Geometry*, 33:247–274, 2005.